

**Theorem 1.** *Assume that we hash  $n$  keys to a table of size  $m$ ,  $n \leq m$ , using a hash function  $h$  chosen at random from a universal family of hash functions, and we resolve collisions by chaining. Then searching for a key takes expected time at most  $1 + \alpha$ .*

**Proof.** For every pair of distinct keys  $k$  and  $l$ ,  $X_{kl} = 1$  iff  $h(k) = h(l)$ , else 0.

$$E[X_{kl}] \leq 1/m.$$

For each key  $k$ , define the r.v.  $Y_k$  that equals the number of keys other than  $k$  that hash to the same slot as  $k$ :

$$Y_k = \sum_{l \in T, l \neq k} X_{kl}$$

$$\text{Thus } E[Y_k] = \sum_{l \in T, l \neq k} E[X_{kl}] \leq \sum_{l \in T, l \neq k} 1/m$$

$$\text{If key } k \notin T: E[Y_k] \leq n/m = \alpha$$

$$\text{If key } k \in T: E[Y_k] \leq (n - 1)/m + 1 < 1 + \alpha \quad \square$$

**Corollary 1.** *Using universal hashing and collision resolution by chaining in a table with  $m$  slots, it takes expected time  $\Theta(s)$  to handle any sequence of  $s$  INSERT, SEARCH, and DELETE operations containing  $O(m)$  INSERT operations.*

# Constructing universal hash functions

Choose a prime number  $p$  such that  $0 \leq k \leq p - 1$ .

Let  $\mathcal{Z}_p = \{0, 1, \dots, p-1\}$  and  $\mathcal{Z}_p^* = \{1, 2, \dots, p-1\}$ .

Let

$$h_{a,b}(k) = ((ak + b) \bmod p) \bmod m$$

be a hash function for any  $a \in \mathcal{Z}_p^*$  and any  $b \in \mathcal{Z}_p$ .

Define the set of hash functions:

$$\mathcal{H}_{p,m} = \{h_{a,b} : a \in \mathcal{Z}_p^*, b \in \mathcal{Z}_p\}$$

Can be stored, evaluated, and chosen in  $O(\log p)$  space/time.

**Theorem 2.**  $\mathcal{H}_{p,m}$  is universal.

**Proof.** Consider two distinct keys  $k$  and  $l$  from  $\mathbb{Z}_p$ . For a given hash function  $h_{a,b}$  we let

$$r = (ak + b) \bmod p$$

$$s = (al + b) \bmod p$$

$r \neq s$  because

$r - s = a(k - l) \pmod{p}$  and both  $a$  and  $k - l$  are nonzero  $\pmod{p}$ .

Moreover, each distinct pair  $(a, b)$  with  $a \neq 0$  yields a distinct pair  $(r, s)$  with  $r \neq s$  because we can solve for  $a$  and  $b$  uniquely given  $r$  and  $s$ :

$$a = ((r - s)((k - l)^{-1} \bmod p)) \bmod p \text{ and}$$

$$b = (r - ak) \bmod p.$$

Thus, for given pair of distinct keys  $k$  and  $l$ , if we pick  $(a, b)$  randomly then  $(r, s)$  is also picked randomly.

Thus the probability that keys  $k$  and  $l$  collide is equal to the probability that  $r = s(\text{mod } m)$  with  $r \neq s$ .

For a given  $r$ , the number of values of  $s$  such that  $r \neq s$  and  $r = s(\text{mod } m)$  is at most  $\lceil p/m \rceil - 1 \leq (p + m - 1)/m - 1 = (p - 1)/m$ .

Thus probability that  $s$  collides with  $r$  is  $\leq 1/m$ .  $\square$

# Perfect Hashing

The **worst case** time to search is  $O(1)$ .

**Definition 1.** *A family  $H$  of hash functions from  $U$  to  $M$  is said to be a **perfect hash family** if for each set  $S \subseteq U$  of size  $|S| \leq |M|$  there exists a hash function  $h \in H$  that is perfect for  $S$ .*

## A two-level hashing scheme

1. Use universal hashing (from the class  $\mathcal{H}_{p,m}$ ) to hash a given set of  $n$  keys to a table of  $m = n$  slots.
2. For  $0 \leq j \leq m - 1$ :  
let  $n_j$  be the number of elements hashed to slot  $j$ ;  
use universal hashing again (from the class  $\mathcal{H}_{p,n_j^2}$ )  
to hash these in a secondary table of size  $n_j^2$ .

**Theorem 3.** *There exists a two-level hashing scheme that uses  $O(n)$  space (for a set of  $n$  keys) and guarantees a worst case search time of  $O(1)$  assuming that the set of keys is **static**.*

**Theorem 4.** *If we store  $k$  keys in a hash table of size  $k^2$  using universal hashing (from the class  $\mathcal{H}_{p,k^2}$ ), then the probability of there being any collision is  $< 1/2$ . Thus, there **exists** an hash function that that will not produce any collision. Moreover, such a function can be found in expected constant number of trials.*

## Proof

$X$  be the number of collisions.

$$E[X] = \binom{k}{2} 1/k^2 < 1/2$$

Markov's inequality gives the result.

**Theorem 5. [Markov Inequality]** *For any non-negative random variable*

$$\Pr(X \geq a) \leq \frac{E[X]}{a}.$$

**Proof.**

$$E[X] = \sum i \Pr(X = i) \geq a \sum_{i \geq a} \Pr(X = i) = a \Pr(X \geq a).$$

□